IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11371

Phishing Attacks Detection System Using Machine Learning

Faisal A. Patel[1], Suraj A. Naphade[2], Kamran A. Shaikh[3], Saurabh V. Phirke[4]

Researcher, Dept. of Computer Engineering, SSBT's COET, Jalgaon, India[1, 2, 3, 4]

Abstract: With the digital revolution around the world more and more number of users are now connecting to the internet, using digital platforms and preferring online or digital banking instead of using cash while making payments. But rise of online transaction have also given opportunity to hackers and fraudsters to fool people and harm them financially. Phishing attacks are type of cyber-crime in which scammers usually send malicious and spam emails, messages and SMS. Some people fall prey to these messages and they contact on the number mentioned or click on the link given in message by this way scammers loot them. The proposed Phishing Attacks Detection System uses Machine Learning Algorithms to identify malicious messages and alert the user. Proposed system uses Naive Bayes algorithm for classification of input data. This will reduce the chances of possible Phishing Attack, identity theft and user will be safe from the financial loss.

Keywords: Phishing Attacks, SMS, E-mail, Machine Learning, Naïve Bayes

I. INTRODUCTION

Now a days people are getting more familiar with online banking and there is a increase in online banking users. But cyber-attacks, phishing and scams are also increasing. people receive some email, SMS or messages from scammers. In those messages there is a link to a fake websites or a phone number to contact with them. some people fall in that trap and their trap and their bank account are emptied in seconds. In recent times machine learning is being used to identify this type of fake messages. but classification of this messages using machine learning is done by some big companies and for the limited usage and specific usage for example Gmail uses classification of emails to check whether they are spam and harmful to the user. This type of systems are still not available to the general public. That's why there is a need of this project so that user can verify any type of message they receive and eventually they will be saved from the scams. website.

People receive hundreds of messages S.M.S and emails every day. With the digital revolution going around people usually prefer digital or online transactions many messages, QR code and links are shared to make online payments some of them are useful and some of them are fraud or phishing attacks. But almost no one has enough amount of time to check authenticity of every message. Clicking on the link given in the messages can lead to a phishing attack and identity theft. So, there is a need of a system which can be check and classify messages and alert user.

A. This system focuses on following parameters:
1. Easy to develop.
2. It is Less Expensive.
3. Large number of users can use proposed system.
4. Detect phishing attempts.
5. Alert the user.

To meet the all this parameters "Phishing Attacks Detection System Using Machine Learning" is proposed.

II. FEATURES FOR SYSTEM REQUIREMENTS

In This section the requirements for "Phishing Attacks Detection System Using Machine Learning" is described. As per described in previous section for parameter (1) and (2) Python programming is used for rapid and easy development of project. For (3) To make project accessible for anyone using any device project will hosted on Web hosting platform. For (4) Machine learning algorithm "Naive Bayes" will be used. For (5) Machine Learning model will classify data and display output.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11371

III. EXISTING SYSTEM

In existing system like spam classification system used in Gmail, it classifies e-mails and puts them in spam folder if they are found containing suspicious links, numbers and some specific keywords.

A. Drawbacks of the Existing System

- Not open source
- Limited only by Gmail to classify e-mails.
- Cannot be used to classify SMS coming from other sources
- Low accuracy

IV. PROPOSED SYSTEM

A. Proposed Approach

The proposed "Phishing Attacks Detection System" is built using various technologies and tool. Dataset is downloaded from Kaggle. Jupyter Notebook is used for data analysis and building Machine Learning Model. Streamlit is a library of python which is used to create simple web page for user interaction. Lastly Heroku platform is used to host the website on internet so everyone can use the application.

V. OBJECTIVE

Objective of the project is to give a solution which will use machine learning algorithm and develop a system which will classify messages, analyze and alert the user if message is looking suspicious and save him from potential phishing attack. The proposed project will be developed using Python language and later it will be deployed to the web host so that anyone connected to the internet can use the application.

This system will provide a place for everyone to check suspicious SMS, emails. Messages from any source can be given as an input and Machine Learning Model will do classification and alert the user if messages are harmful.

VI. IMPLEMENTATION

Phishing Attacks Detection system is developed by using Python 3.9. Firstly, dataset is downloaded from Kaggle. To perform Machine Learning tasks important libraries of python are installed in system.
Following are the libraries required for the development of the system.

1. NumPy
2. Pandas
3. Matplotlib
4. Streamlit
5. Nltk
6. Sklearn
7. Pickle

Data analysis and preprocessing is performed using jupyter notebook in following steps.

1. Data Cleaning
2. EDA
3. Text Preprocessing
4. Model Building
5. Evaluation
6. Improvement

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11371

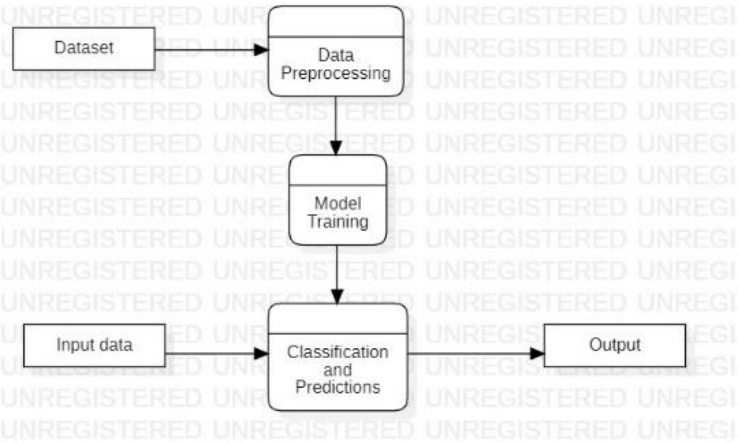

Fig. 1 System Architecture

VII. RESULT AND DISCUSSION

The proposed system is developed in Python programming language. To avoid any future risks due to update in the functionality of python code used, Virtual Environment is used. The algorithm used for classification is Naive Bayes. As the project input data is in text form Naive Bayes performs better than other machine learning algorithm. There are multiple types of Naive Bayes algorithm, mainly Bernoulli Naive Bayes, Multinomial Naive Bayes and Gaussian Naive Bayes. All of this algorithm were given the training data, tested on testing data and their accuracy and precision was observed.

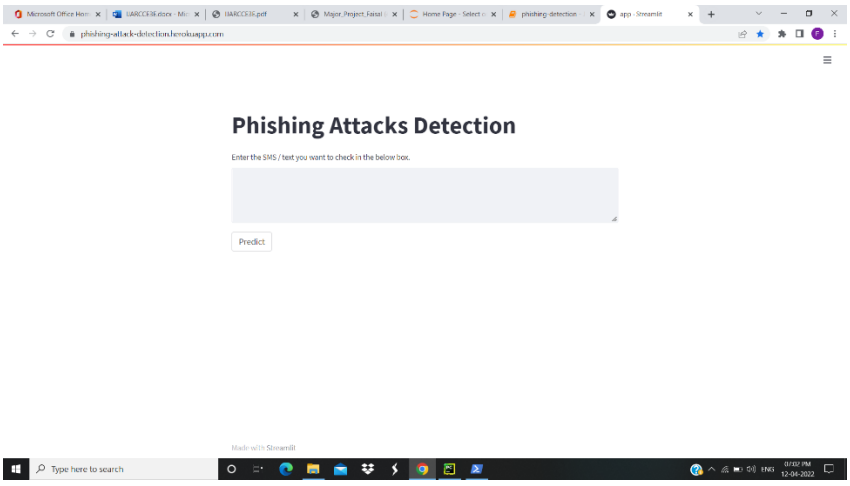

Fig. 2 Home Screen of the proposed system.

Gaussian Naive Bayes gave accuracy of 0.86 and precision of 0.50, Bernoulli Naive Bayes gave accuracy of 0.98 and precision of 0.99, Multinomial Naive Bayes gave accuracy of 0.97 and precision 1.0. After observing accuracy and precision of each of this algorithms Multinomial Naïve Bayes is selected to be used in the Phishing Attacks Detection System.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11371

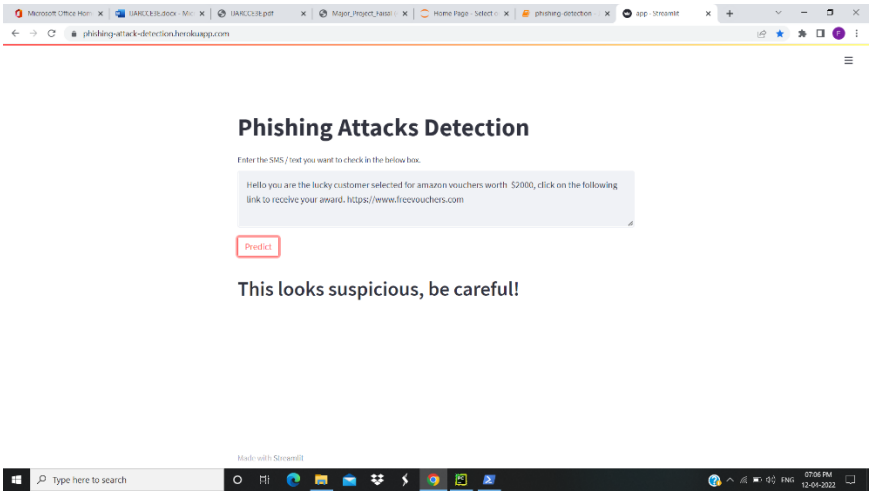

Fig. 3 An alert displayed if harmful message is detected.

VIII. CONCLUSION

Phishing Attacks Detection system provides a place for people to check any suspicious looking message, email and SMS which can possibly lead to phishing attack. The proposed system uses well-known classification algorithm Naive Bayes. The observed comparative analysis of different algorithms has different levels of accuracy to determine the effectiveness and efficiency of predictions. Compared to previously available system the Machine Learning model provides higher accuracy of detecting phishing attacks. This will save users from phishing attacks, financial loss, identity theft and data breach.

The Phishing Attacks Detection system met with all the requirements described in user requirements. Phishing Attacks Detection system can be used by anyone using any device as it is hosted on website. The system fulfils its objective and performs better than the existing systems.

REFERENCES

[1]. APWG organization," Phishing Activity Trends Reports", https://apwg.org/ trendsreports/
[2]. Ram Basnet, Srinivas Mukkumala, Andrew H. Sung," Detection of Phishing Attacks: A Machine Learning Approach".
[3]. Zulfikar Ramzan," Phishing Attacks and Countermeasures", https://link.springer. com/chapter/10.1007/978-3-642-04117-4_23
[4]. Tiago A Almeida, Jos Mara, G. Hidalgo," Contribution to the study of SMS spam filtering", https://dl.acm.org/doi/10.1145/2034691.2034742
[5]. Dataset from Kaggle," SMS Spam Collection Dataset", https://www.kaggle.com/ uciml/sms-spam-collection-dataset
[6]. Analytics India Magazine, "7 Types of Classification Algorithms", https:// analyticsindiamag.com/7-types-classification-algorithms/

BIOGRAPHIES

Mr. Nitin P. Jagtap, completed B.E (IT) and M.E. in Computer Science & Engineering. He is working as Assistant Professor in SSBT's College of Engineering and Technology since 2007. He is pursuing his PhD in Computer Science & Engineering in Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. Project Guide for "Phishing Attacks Detection System Using Machine Learning". Area of Interest: Data mining, Machine Learning, Sentiment Analysis, and Data Analytics.

Mr. Faisal A. Patel, Researcher, Dept. of Computer Engineering, Shrama Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon. Area of Interest: Machine Learning, Python programming, Data analysis.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11371

Mr. Suraj A. Naphade, Researcher, Dept. of Computer Engineering, Shrama Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon. Area of Interest: C, C++, Core Java.

Mr. Kamran A. Shaikh, Researcher, Dept. of Computer Engineering, Shrama Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon. Area of Interest: Web Development, Database Management.

Saurabh V. Phirke, Researcher, Dept. of Computer Engineering, Shrama Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon. Area of Interest: C, C++, Web development.